JHU-NJU Survival Analysis

Lab 2 (July 19, 2011)

1 PBC Data

NAME: PBC Data (PBC.DAT) SIZE: 418 observations, 20 variables

SOURCE: Counting Processes and Survival Analysis by T. Fleming, D. Harrington, (1991), published by John Wiley and Sons.

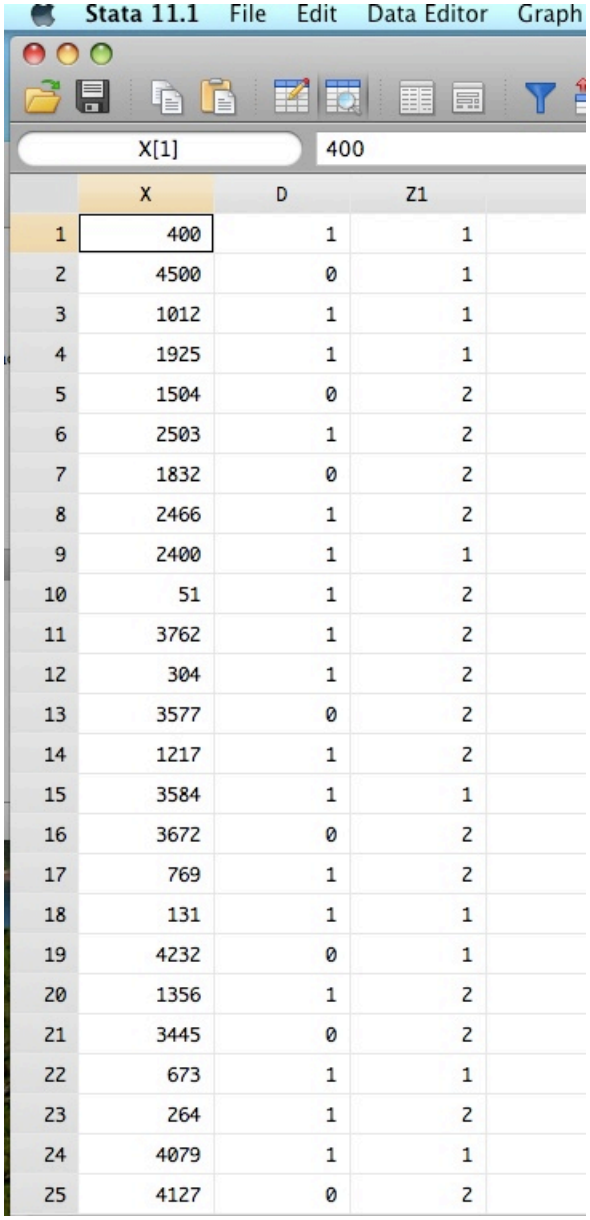
BASIC DATA DESCRIPTION: Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. Censoring was due to liver transplantation.

The Data

X The number of days between registration and the earlier of death, liver transplantation, or study analysis time in July, 1986.

D 1 if X is time to death, 0 if time to censoring

Z1 Treatment Code, 1 = D-penicillamine, 2 = placebo.



	X	D	Z1
1	400	1	1
2	4500	0	1
3	1012	1	1
4	1925	1	1
5	1504	0	2
6	2503	1	2
7	1832	0	2
8	2466	1	2
9	2400	1	1
10	51	1	2
11	3762	1	2
12	304	1	2
13	3577	0	2
14	1217	1	2
15	3584	1	1
16	3672	0	2
17	769	1	2
18	131	1	1
19	4232	0	1
20	1356	1	2
21	3445	0	2
22	673	1	1
23	264	1	2
24	4079	1	1
25	4127	0	2

2 Comparing treatment groups

We know the MLE (maximum likelihood estimate) for θ in $T \sim \exp(\theta)$ is $\hat{\theta} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} y_i}$.

Performing this operation, we can obtain the MLEs for both treatment groups :

Intervention (group 1): $\hat{\theta}_1 = .0002041$

Placebo (group 2): $\hat{\theta}_2 = .0001951$

- What are the assumptions of this model?
- How is the MLE related to how we would usually estimate the failure rate?

We can also find the variance of these estimates using Fisher's Information:

$$Var(\hat{\theta}) = \frac{\hat{\theta}^2}{\sum_{i=1}^{n} \delta_i} \tag{1}$$

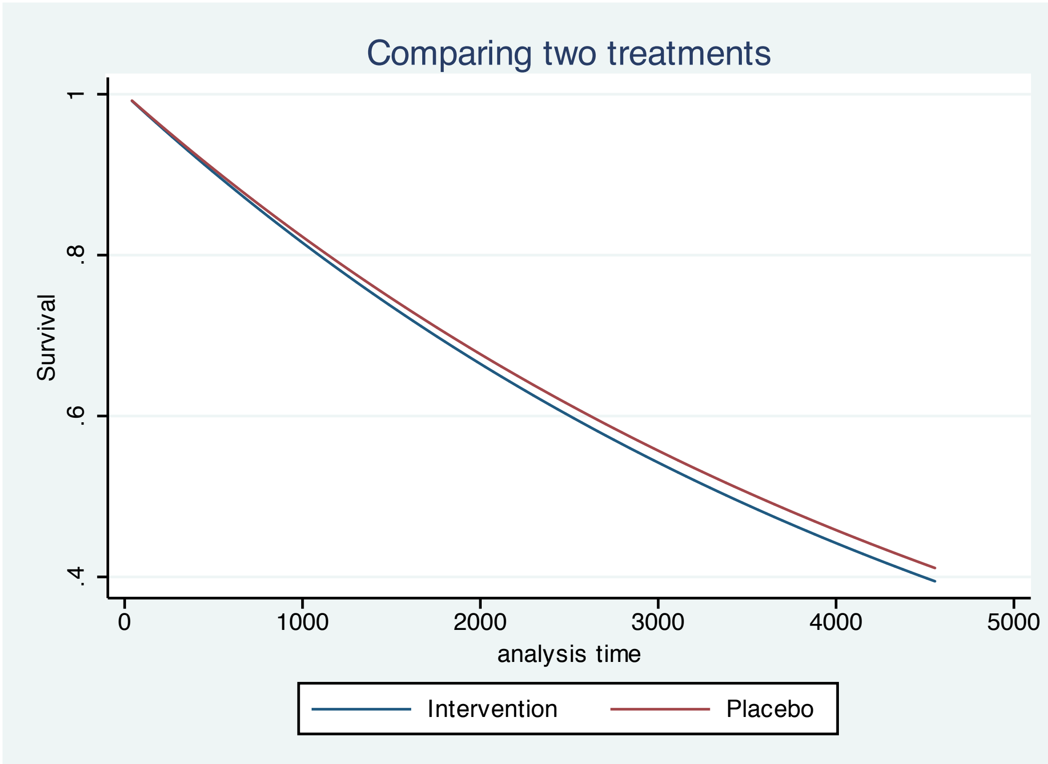
Using the variance, we can find the 95% confidence interval for the MLE's:

$$CI_{95} = (\hat{\theta} - 1.96 \cdot SE_{\hat{\theta}}, \hat{\theta} + 1.96 \cdot SE_{\hat{\theta}})$$

Using this, we have the following confidence intervals:

Group 1: $CI_{95} = (0.0001545, 0.0002537)$

Group 2: $CI_{95} = (0.0001457, 0.0002445)$



3 Appendix: STATA Code

```
*generate summation of censoring variable and time variable for group 1

egen sumD=sum(D) if Z1==1

egen sumX=sum(X) if Z1==1

*find the MLE for group 1

gen mle1=sumD/sumX if Z1==1


*generate summation of censoring variable and time variable for group 2

egen sumD2=sum(D) if Z1==2

egen sumX2=sum(X) if Z1==2

*find the MLE for group 2

gen mle2=sumD2/sumX2 if Z1==2
```

```
*code from Prof. Wang

*set survival data

stset X, failure(D)


*recode variables for later functions

gen z01=Z1

replace z01=0 if Z1==2


*apply exponential distribution

streg z01, d(exponential) nohr

predictnl haz = predict(hazard), ci(haz_lb haz_ub)

*plot estimates

stcurve, survival at1(z01=1) at2(z01=0)
```